Analytics Design to Recommend Flavors to Launch: Part II

You should reread the workshop case **Analytics Design to Recommend Flavors to Launch: Part I**. That case gives some important background and data description of the system data.

To briefly summarize, you are conducting analytics to provide input or recommendations to the decision of which flavors to launch next. Assume for this analysis that the private label recently launched six flavors--Blueberry, Honey, Peach, Plain, Strawberry, and Vanilla. You should focus on providing analytics to help determine what the next flavors to launch should be.

You now have access to the data from a survey that captures information about consumers' flavor preferences. The details of the survey are provided below. You also now have access to data pulled from the loyalty card system that the data architect created for the project. Access is possible through a MySQL database (see blackboard site for instructions).

During conversations with the VP and private brand manager, the analytics goals have become clearer. In particular, the VP laid out that your team will deliver analyses related to the following loose analysis design.

The design has four measurement goals:
1) Describe the percentage of sales of existing flavors in the Greek yogurt category (all brands)
2) Describe the percentage of sales of existing yogurt flavors outside of Greek yogurt (regular class of yogurt)
3) Describe survey respondents' preferences for Greek yogurt flavors
4) Predict the best set of next flavors to add to achieve the highest reach using the survey data.

You need to refine this design to be more precise (exact data requirements, exact analyses), implement the design, and create a presentation that delivers on the goal of providing input and recommendations on the next flavors to launch. The presentation should be focused on the VP and private brand manager. Both are not particularly comfortable with statistical analysis, but quite comfortable with quantitative information. They care most about seeing the results on items 1-4 above along with a summary of the key points and recommendations. You should integrate the findings, not merely present each one.

This assignment should contain
1 Title slide
1 Highlights slide
3-4 Detail slides
1 methodology slide (optional to describe your predictive methodology)
An .Rmd file as appendix

The .Rmd file should be a clean file containing up to three parts:
1) Code to produce the output in slides. This should be clearly commented to communicate what it does and why. Each slide should have its own separate R code block and be clearly marked with "OUTPUT FOR SLIDE N", where N is the slide number, and surrounded by a line of ########### symbols. You can add comments and discussion in the text surrounding the code blocks to explain your choices if you so desire. This comments should be concise!
2) After or before the main code blocks in (1), a block of code that represents the SQL statements used to pull the data used in your assignment. This should be demarked with "SQL CODE" and surrounded by a line of ############ symbols.
3) Any additional relevant analyses you do that do not appear in the slides themselves can be put at the very bottom of the .Rmd file. These analyses should be clearly commented to explain what you did and why. This group of code blocks should be marked with "ADDITIONAL ANALYSES" and surrounded by a line of ############ symbols.

Evaluation

Frame:

- Refined analysis design meets decision needs (3 pts)
- Analysis design addresses major challenges in data and analysis (1 pt)

Analysis:

- Analytics/queries appear to be implemented correctly (2 pts)
- Information and insight presented provides key points from data (2 pts)

Communicate: (2 pts)

- Main points are easy to identify in titles and subtitles
- Tables/figures annotated, axes labeled, etc.
- Use of footnotes provides proper sourcing without too much distraction
- Communication appropriate for audience (language, use of statistics)
- Use of space and visual highlighting aligns with key messages

Survey Data

The survey was taken as a non-random sample from the population of customers. The invitation for the survey went to those that had purchased Greek Yogurt in the last 30 days. This screening criteria was calculated based on household, not individual level data. The survey was pre-tested the week prior to the survey being sent out. It was sent out on May 15, 2011. Some surveys were incomplete and some took an unusually long time to complete, which the panel team usually considered a bad sign for the quality of response. Some decided to skip some questions (for Q12, this is identified by all 23 flavors being NA).

The survey data are described below. The survey data consist of three tables. One table, survQuestions, contains the survey question text as a single row. The second, survResponses, has the same columns and contains the respondents' responses. The third, survItemSales, contains all of the relevant item purchases related to the users in the survey. This third table can be located in the system data.

1. ID: *The UserID for the respondent.*

2. V8: Start Date/Time of the response

3. V9: End Date/Time of the response

4. V10: Whether finished the response
Values: 1= finished, 0=not finished

5. Q1: Have you eaten Greek Yogurt in the past month?
Radio Button: Yes (1) or No (0). NA means no response.

6. Q2: Of all the yogurt you have eaten in the past month, what percentage was Greek Yogurt?
Slider: Value between 0 and 100, NA means no response.

7-14. Q3: What attracts you to Greek Yogurt? (Check all that apply)
Multiselect Checkbox: Value of 1 for each column they checked otherwise blank (NA)

15-37. Q12: Below is a list of both existing and new flavors. Assume these flavors are all available for your consumption. Please organize the flavors below into how often you would eat each one.
Each column corresponds to a flavor and contains a 0 (Regularly), 1 (Occasionally), 2 (Never) or NA (did not categorize that flavor or did not complete the question).